APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/421,416	10/19/1999	HINRICH SCHUETZE	D/99198	4124

7590 01/24/2003

JOHN E BECK
XEROX CORPORATION
XEROX SQUARE 20A
ROCHESTER, NY 14644

EXAMINER
FLEURANTIN, JEAN B

ART UNIT	PAPER NUMBER
2172	12

DATE MAILED: 01/24/2003

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 07-01)

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) FROM
THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed
  after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any
  earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _CPA, 11/14/02_ .

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is
closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1,7-39,41-43 and 45-49_ is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1,7-22,28,39,41-43,45,47 and 49_ is/are rejected.

7)☒ Claim(s) _23-27,29-38,46 and 48_ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

11)☐ The proposed drawing correction filed on _____ is: a)☐ approved b)☐ disapproved by the Examiner.

    If approved, corrected drawings are required in reply to this Office action.

12)☐ The oath or declaration is objected to by the Examiner.

Priority under 35 U.S.C. §§ 119 and 120

13)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____ .

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage
application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

14)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).

    a) ☐ The translation of the foreign language provisional application has been received.

15)☐ Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449) Paper No(s) _____ .

4)☐ Interview Summary (PTO-413) Paper No(s). _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: .

DETAILED ACTION

1.   Claims 7-39, 41-43 and 45-49 are remained pending for examination.

Continued Prosecution Application

2.   The request filed on 11/04/02 for a Continued Prosecution Application (CPA) under 37

CFR 1.53(d) based on parent Application No. 09/421,416 is acceptable and a CPA has been

established. An action on the CPA follows.

Claim Rejections - 35 U.S.C. § 103

3.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in
> section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are
> such that the subject matter as a whole would have been obvious at the time the invention was made to a person
> having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the
> manner in which the invention was made.

Claims 7-22, 28, 39, 41-43, 45, 47 and 49 are rejected under 35 U.S.C. 103(a) as being

unpatentable over Tanaka et al. (US Pat. No. 5,684,891) in view of Pirolli et al. (US Pat. No.

5,835,905), ("Tanaka"), ("Pirolli").

As per claims 1 and 39, Tanaka teaches a method for quantitatively representing

documents in a vector space (see col. 10, lines 47-49), as claimed comprises the steps of

identifying a first document to be processed from a plurality of objects documents (thus, the

character images and requests the general classifier 25 to perform general classification

processing after feature extraction of the first character of the received image ends; which is

readable as identifying a first document to be processed from a plurality of objects documents)

(see col. 14, lines 45-49). Further, in column 15, lines 1 through 3, Tanaka teaches detailed

identification processing are executed by the character recognition board 4 using the unprocessed

feature vectors left in the ram 23;

extracting a first feature corresponding to the first document from the plurality of

documents(thus, the central processing unit 21 executes feature extraction of the character

images from the second character onward; which is readable as extracting a first feature

corresponding to the first document from the plurality of documents) (see col. 14, lines 51-52),

the first feature comprising text surrounding an image included in the document, the text

surrounding the image not being anchor text (thus, recognition is applied to laterally written text,

such as a Japanese sentence in which a circumscribed rectangle of a character image of each

character constituting the sentence approximates a square and the character pitch is fixed; which

is readable as text surrounding an image included in the document, the text surrounding the

image not being anchor text) (see col. 15, lines 49-53);

converting the first feature to a first vector (thus, the processing proceeds to step S43 at

which an interrupt is generated in the central processing unit 11 and an image data transfer

request instruction is issued in order to request the next character image, at the same time the

general classifier 25 is instructed at step S44 to perform matching between the obtained feature

vector; which is readable as converting the first feature to a first vector) (see col. 13, lines 23-29).

But, Tanaka does not explicitly indicate steps of associating the first vector with the first

document. However, Pirolli implicitly indicates feature vector representations of each web page

that represent the value of each page on each dimension and which are used in the categorization process, (see col. 6, lines 1-3). Thus, it would have been obvious to a person of ordinary skill in the art at the time the invention was made to modify the teachings of Tanaka and Pirolli with steps of associating the first vector with the first document. This modification would allow the teachings of Tanaka and Pirolli to improve the accuracy and the reliability of the system and method for quantitatively representing data objects in vector space, and provide facilitate information searching (see cols. 1-2, lines 67-1).

As per claims 7, 10 and 15, Tanaka teaches a method as claimed, further comprises the steps of extracting a second feature corresponding to the document (thus, the features are extracted from the character images of the character rectangles 351 at step S503, which is readable as extracting a second feature corresponding to the document) (see col. 19, lines 45-46),

converting the second feature to a second vector (thus, the flag in the buffer of RAM 23 which stores the feature vector is investigated and it is examined whether the general classifier 25 is executing processing, if it is determined that processing is in progress then the completion of general classification processing is awaited, when this has been completed processing proceeds to step S50, at which general classification processing base upon a new feature vector is requested of the general classifier 25; which is readable as converting the second feature to a second vector) (see col. 14, lines 1-9). But, Tanaka does not explicitly indicate URL representing the first document; and associating the second vector with the first document. However, Pirolli indicates various features of the pages such as file size and URL; and feature

vector representations of each web page that represent the value of each page on each dimension

and which are used in the categorization process, (see cols. 5 and 6, lines 58-60 and 1-3). Thus,

it would have been obvious to a person of ordinary skill in the art at the time the invention was

made to modify the teachings of Tanaka and Pirolli with URL representing the first document;

and associating the second vector with the first document. This modification would allow the

teachings of Tanaka and Pirolli to improve the accuracy and the reliability of the system and

method for quantitatively representing data objects in vector space, and provide facilitate

information searching (see cols. 1-2, lines 67-1).

As per claims 8, 11, 13, 16, 18, 19 and 41, the limitations of claims 8, 11, 13, 16, 18, 19

and 41 are rejected in the analysis of claim 7, and these are rejected on that basis.

As per claims 9 and 14, the limitations of claims 9 and 14 are rejected in the analysis of

claim 7, and these are rejected on that basis.

As per claims 12, 17 and 42, Tanaka teaches the claimed subject matter except the

claimed, wherein the numeric value representative of the number of links in each corresponding

document linking to the document is calculated as the token frequency weight of the

corresponding link multiplied by the inverse context frequency weight of the corresponding link.

However, Pirolli teaches token statistics for the web page are then generated, these statistics

include token occurrence the token information is then used to create a document vector, where

each component of the vector represents a word, entries in the vector for a document indicate the

presence or frequency of a word in the document, (see col. 7, lines 55-61). Thus, it would have

been obvious to a person of ordinary skill in the art at the time the invention was made to modify

the teachings of Tanaka and Pirolli with wherein the numeric value representative of the number

of links in each corresponding document linking to the document is calculated as the token

frequency weight of the corresponding link multiplied by the inverse context frequency weight of

the corresponding link. This modification would allow the teachings of Tanaka and Pirolli to

improve the accuracy and the reliability of the system and method for quantitatively representing

data objects in vector space, and provide facilitate information searching (see cols. 1-2, lines 67-

1).

As per claim 20, Tanaka teaches a method as claimed, wherein the second feature

comprises a text genre feature (see col. 11, lines 24-26).

As per claim 21, Tanaka teaches the claimed subject matter except the claimed,

processing the first to calculate the probability that the first document is of the corresponding text

genre; and creating the second vector having a number of dimension equal to the number of

possible text genres, and the second vector further having as each element a numeric value

representative of the probability that the first document is of the corresponding genre. However,

Pirolli teaches techniques from information retrieval can be applied to calculate a text similarity

matrix which represents the inter-document text similarities among web pages, in particular for

each web page the text is tokenized and indexed using a statistical content analysis process; token

statistics for the web page are then generated, these statistics include token occurrence the token

information is then used to create a document vector, where each component of the vector

represents a word, entries in the vector for a document indicate the presence or frequency of a

word in the document (see col. 7, lines 35-39, 55-61). Thus, it would have been obvious to a

person of ordinary skill in the art at the time the invention was made to modify the teachings of

Tanaka and Pirolli with processing the first to calculate the probability that the first document is

of the corresponding text genre; and creating the second vector having a number of dimension

equal to the number of possible text genres, and the second vector further having as each element

a numeric value representative of the probability that the first document is of the corresponding

genre. This modification would allow the teachings of Tanaka and Pirolli to improve the

accuracy and the reliability of the system and method for quantitatively representing data objects

in vector space.

As per claims 22, 28, 45 and 47, Tanaka teaches a method as claimed, wherein the first

feature comprises the color histogram for the image included in the first document (thus, a

histogram of black pixels in the horizontal direction may be taken for every rectangle and a line

may be drawn in the portion where the histogram is maximum, which is equivalent to wherein

the first feature comprises the color histogram for the image included in the first document) (see

col. 27, lines 5-21).

As per claims 43 and 49, in addition to the discussion in claims 1 and 7, Tanaka further

teaches selecting an image feature as a first feature, the image feature being associated with the

non-text content of an image included in the document; and a second a second feature from a set

multi-modal features including a user information feature and a genre feature (thus, selection

means for evaluating results of character recognition performed by the character recognition

means and selecting as a final segmented character image, a segmented character image having

the highest evaluation among the several segmented character images segmented by the second

segmentation means; which is readable as selecting an image feature as a first feature, the image

feature being associated with the non-text content of an image included in the document; and a

second a second feature from a set multi-modal features including a user information feature and

a genre feature) (see col. 7, lines 42-48).

As per claim 44, the limitations of claim 44 are rejected in the analysis of claim 22, and

this is rejected on that basis.

Allowable Subject Matter

4.   Claims 23-27, 29-38, 46 and 48 are objected to as being dependent upon a rejected base

claim, but would be allowable if rewritten in independent form including all of the limitations of

the base claim and any intervening claims.

5.   The prior art made of record and not relied upon is considered pertinent to applicant's

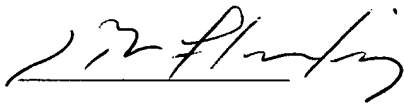disclosure: Pirolli et al. US Pat. No. 5,895,470 relates to design linked collections of documents.

Conclusion

6.   Any inquiry concerning this communication from examiner should be directed to Jean

Bolte Fleurantin at (703) 308-6718. The examiner can normally be reached on Monday through

Friday from 7:30 A.M. to 6:00 P.M.

If any attempt to reach the examiner by telephone is unsuccessful, the examiner's

supervisor, Mrs. KIM VU can be reached at **(703) 305-8449.** The FAX phone numbers for the

Group 2100 Customer Service Center are: *After Final* **(703) 746-7238,** *Official* **(703) 746-7239,**

and *Non-Official (703) 746-7240.* NOTE: Documents transmitted by facsimile will be entered

as official documents on the file wrapper unless clearly marked *"DRAFT"*.

Any inquiry of a general nature or relating to the status of this application or proceeding

should be directed to the Group 2100 Customer Service Center receptionist whose telephone

numbers are **(703) 306-5631, (703) 306-5632, (703) 306-5633.**

Jean Bolte Fleurantin

January 22, 2003

JBF/

JEAN M. CORRIELUS
PRIMARY EXAMINER